



Audiovisual integration of stimulus transients

Andersen, Tobias; Mamassian, Pascal

Published in:
Vision Research

Link to article, DOI:
[10.1016/j.visres.2008.08.018](https://doi.org/10.1016/j.visres.2008.08.018)

Publication date:
2008

Document Version
Early version, also known as pre-print

[Link back to DTU Orbit](#)

Citation (APA):
Andersen, T., & Mamassian, P. (2008). Audiovisual integration of stimulus transients. *Vision Research*, 48(25), 2537-2544. <https://doi.org/10.1016/j.visres.2008.08.018>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Audiovisual integration of stimulus transients

Tobias S. Andersen^{a,b,*}, Pascal Mamassian^c

^a Center for Computational Cognitive Modeling, Department of Psychology, University of Copenhagen, Linnésgade 22, 1361 Kbh. K., Denmark

^b Department of Informatics and Mathematical Modeling, Technical University of Denmark, Richard Petersens Plads, Building 321, 2800 Lyngby, Denmark

^c Laboratoire Psychologie de la Perception, Université Paris Descartes/CNRS UMR 8158, 45 rue des Saints Pères, 75270 Paris Cedex 06, France

ARTICLE INFO

Article history:

Received 4 July 2008

Received in revised form 13 August 2008

Available online xxxx

Keywords:

Multisensory

Vision

Audition

Signal detection

ABSTRACT

A change in sound intensity can facilitate luminance change detection. We found that this effect did not depend on whether sound intensity and luminance increased or decreased. In contrast, luminance identification was strongly influenced by the congruence of luminance and sound intensity change leaving only unsigned stimulus transients as the basis for audiovisual integration. Facilitation of luminance detection occurred even with varying audiovisual stimulus onset asynchrony and even when the sound lagged behind the luminance change by 75 ms supporting the interpretation that perceptual integration rather than a reduction of temporal uncertainty or effects of attention caused the effect.

© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

Multisensory integration often occurs when perceiving environmental attributes such as position, time and phonetic content of speech that are mediated by both sound and light. Multisensory integration is characterized by two perceptual phenomena. First, *crossmodal enhancement* occurs when accuracy of perception is increased by congruent information reaching more than one sensory modality. Second, *crossmodal illusions* can occur when incongruent information reaches different sensory modalities. Many examples of such multisensory illusions are known. These include ventriloquism (Warren, 1979) where the perceived position of a sound source is influenced by a concurrent light and the McGurk effect (McGurk & MacDonald, 1976) where watching a talker's face can change the phoneme perceived from the voice.

Abrupt change, as when an object falls to the floor and breaks or a prey in hiding suddenly sets off in escape, is an environmental attribute that is often mediated by both sound and light. Therefore, we hypothesize that multisensory integration may occur when perceiving auditory and visual intensity transients. More specifically, we expect that an auditory transient can influence the perceived saliency of a visual transient and vice versa. This effect has not been studied directly although some audiovisual perceptual phenomena may depend upon it. Auditory flutter rate influence perceived visual flicker rate (Recanzone, 2003). Perception of flicker and flutter rate is likely to be based on the perception

of intensity transients and if audiovisual integration of transients causes illusory transients or illusory elimination of transients it might underlie this phenomenon. In a similar perceptual phenomenon, the number of perceived flashes is influenced by the number of concurrently presented beeps (Andersen, Tiippana, & Sams, 2004; Shams, Kamitani, & Shimojo, 2000). In the stream/bounce illusion (Sekuler, Sekuler, & Lau, 1997), two dots move in a way that could be interpreted as them streaming through each other or bouncing off one another. When accompanied by a short sound burst at the moment when they overlap, observers are more likely to perceive the dots to bounce rather than to stream. If the dots bounce, there would be a transient in their path, but if they stream, then their path would be unchanging. If the sound induces a percept of a visual transient, this might favor the percept of the dots bouncing.

The aim of the current study is to examine directly whether sound can influence the perception of a visual stimulus transient in a signal detection task. Since Green and Swets' work (Green & Swets, 1966) it has been clear that two weak signals may combine to enhance sensitivity in signal detection. Recently, Schnupp, Dawe, & Pollack (2005) showed that this enhancement occurs across sensory modalities so that it is far easier to detect weak co-occurring auditory and visual signals than it is to detect either alone. This is not surprising as there is indeed more signal energy to detect when there are two signals although the energy is divided in two channels. This crossmodal effect can therefore be due to a change in the response criterion so that the perceptual decision of whether a signal was present or not is based on the Euclidian sum of the auditory and visual internal representations rather than on either one alone. The internal representations of the auditory and visual signals can thus still be independent and accessible de-

* Corresponding author. Address: Department of Informatics and Mathematical Modeling, Technical University of Denmark, Richard Petersens Plads, Building 321, 2800 Lyngby, Denmark. Fax: +45 4588 2673.

E-mail address: ta@imm.dtu.dk (T.S. Andersen).

spite an increased sensitivity when they are combined. That is, if observers were to detect only the auditory signal while ignoring the visual signal or *vice versa*, multisensory integration might not occur. We refer to this type of integration as *combinatory* because it requires the combination of auditory and visual independent internal representations. Combinatory enhancement may underlie crossmodal enhancement but not crossmodal illusions because it leaves the veridical unisensory representations available.

In contrast, *auxiliary* multisensory integration occurs even when the observer tries to report solely on the basis of the relevant modality while ignoring another irrelevant, or auxiliary, modality. Auxiliary integration can thus underlie both crossmodal enhancement and crossmodal illusions. It is this type of multisensory integration that we will investigate in this study. It was only recently that Frassinetti and coworkers systematically demonstrated that such *auxiliary* multisensory integration occurs in signal detection (Frassinetti, Bolognini, & Ladavas, 2002). In their study, they found an enhancement of visual detection sensitivity by sound for co-localized and simultaneous auditory and visual stimuli. Notably, this effect occurred when observers responded only to the visual stimulus while trying to ignore the auditory stimulus.

To study auxiliary audiovisual integration of intensity transients, we designed a 2-Interval Forced Choice (2IFC) signal detection paradigm in which observers were to detect in which interval a luminance change occurred. The luminance change could be either an increase or a decrease. In both intervals, a perceptually co-localized sound increased, decreased or remained constant in intensity. The stimuli are illustrated in Fig. 1. Since the sound in the two intervals was the same, it could not bias observers to respond in one interval or the other. Therefore, the *proportion correct* is a direct measure of visual detection sensitivity and any change in proportion correct must be due to an auxiliary effect of sound on visual detection sensitivity. On each trial observers also identified the luminance as an increase or a decrease. This identification task was thus performed in a yes-/no-paradigm and so the proportion correct is not a direct measure of sensitivity but is prone to response bias and any such bias may be influenced by sound.

If abrupt change *per se* is a stimulus feature that is integrated across modalities, any abrupt acoustic change should facilitate visual change detection. This should lead to an increase in proportion correct in the detection task when the sound intensity changed and this would not depend on whether sound intensity and luminance

increased or decreased. We call this the *transient hypothesis*, because it describes multisensory integration as based on unsigned signal transients. According to the transient hypothesis there should be no effect of sound intensity change in the identification task in which performance must be based on the direction of the change and thus on the sustained luminance level following the transient. These predictions of the transient hypothesis are illustrated in Fig. 2A.

Alternatively, multisensory integration could be between loudness and brightness. We call this the *loudness/brightness hypothesis*. This hypothesis was proposed by Stein, London, Wilkinson, & Price (1996). In their study, observers rated the brightness of a target LED either accompanied by a loud sound or not. Stein et al. found that participants rated the brightness of the LED higher when it was accompanied by the sound. If this effect persists near visual threshold, then a loud sound making a weak visual stimulus appear brighter will facilitate detection of the visual stimulus. However, if the task is to detect a luminance decrease, then a loud sound at the time of the decrease should make the luminance appear brighter and therefore make the decrease appear smaller and thus impede detection. In the identification task, the loudness/brightness hypothesis predicts that a sound intensity increase will make the luminance change appear more as an increase regardless of whether it was in fact an increase or a decrease. Of course, the opposite goes for a sound intensity decrease. These predictions of the loudness/brightness hypothesis are illustrated in Fig. 2B.

Odgaard, Arie, & Marks (2003) pointed out that Stein et al.'s results might be confounded by response bias effects. The loud sound could simply increase observers' tendency to report the light brighter than the intensity they actually perceived. In much the same way, an increasing tendency to report that a visual signal was present when accompanied by a sound does not reflect observers' increased sensitivity to that signal but simply a change in response bias. Such response bias effects were consistently found by Frassinetti et al. in addition to the perceptual effects described above. When Odgaard et al. used a paradigm where the effect of response bias was constant, they found no interaction between loudness and perceived brightness. Thus, according to the *bias hypothesis*, the direction of the sound intensity change bias observers' responses in the luminance identification task in the same direction. However, a response bias will not affect performance in the detection task. The predictions of the bias hypothesis

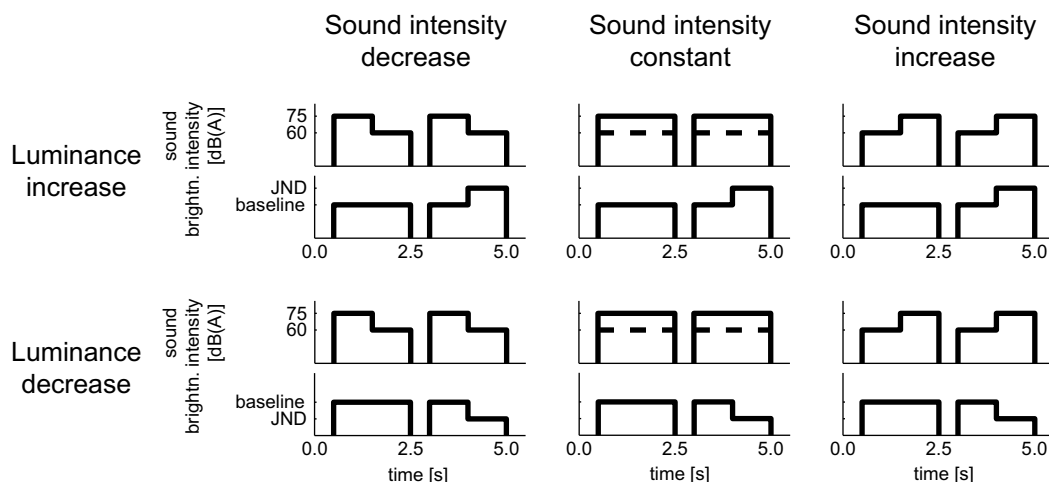


Fig. 1. The sound intensity and brightness as a function of time for each of the six trial types in Experiment 1. A trial contained two stimulus intervals. In both intervals, sound intensity decreased from 75 dB(A) to 60 dB(A), increased from 60 dB(A) to 75 dB(A) or remained constant. When the sound intensity was constant it was either 75 dB(A) or 60 dB(A) (indicated by dashed line). In one interval, the brightness (brightn.) either increased or decreased from a baseline value that was adjusted to each participant to the approximate Just Noticeable Difference (JND). In the other interval, brightness remained constant. The order of the intervals varied pseudorandomly between trials.

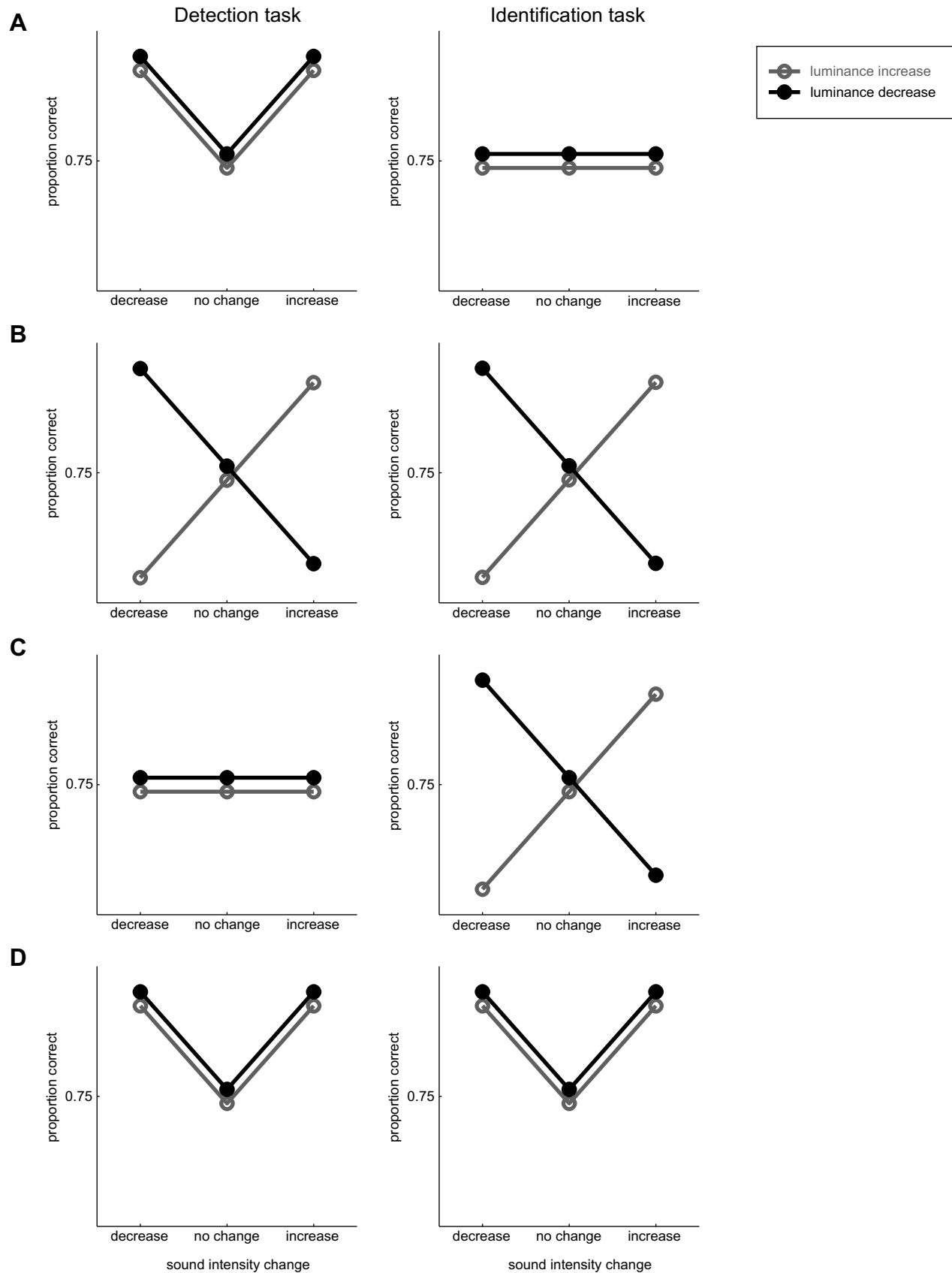


Fig. 2. Predictions of the transient (A), loudness/brightness (B), bias (C) and attention/uncertainty (D) hypotheses for Experiment 1. The predictions are overlapping for luminance increase and decrease in the absence of an interaction between luminance and sound intensity. Therefore a small vertical shift is introduced between black and grey lines for illustrative purposes.

in the absence of any perceptual effects are displayed in Fig. 2C. We note that bias effects are not exclusive and might occur concurrently with other effects.

Perceptual effects can be difficult to separate from attentional effects. According to the *attention hypothesis*, the interaction between auditory and visual signals is due to the signal in one modality cueing attention to the occurrence of the signal in the other modality. A salient stimulus draws exogenous, involuntary attention not only to its location (Posner, 1980; Posner, Snyder, & Davidson, 1980) but also to the time of its occurrence (Nobre & O'Reilly, 2004). Spatial attention works across sensory modalities (Spence & Driver, 1997) so that an irrelevant sound can increase visual sensitivity (McDonald, Teder-Salejari, & Hillyard, 2000). Thus, cross-modal attentional cueing share many of the features of crossmodal sensory integration and it is controversial whether the two effects are separable at all or whether they are in fact maintained by the same system (Macaluso, Frith, & Driver, 2000; McDonald, Teder-Salejari, & Ward, 2001). A related effect is that of reduction of uncertainty. If the signal in one modality provides information on the time of occurrence of the signal in the other modality then temporal uncertainty is reduced. This reduces the temporal interval that the observer needs to monitor for the occurrence of a target and will thus effectively increase the signal-to-noise ratio which will result in an increased sensitivity to the target stimulus (Pelli, 1985). We expect an effect of the sound intensity change cueing attention or reducing uncertainty to be one of a general increase in the proportion correct in both the detection and identification tasks as illustrated in Fig. 2D.

In our second experiment, we varied the crossmodal Stimulus Onset Asynchrony (SOA) to study the temporal dynamics of audiovisual integration. Many multisensory effects only occur when the auditory and visual stimuli coincide within a temporal window of approximately 100 ms (Meredith, Nemitz, & Stein, 1987; Shams, Kamitani, & Shimojo, 2002). Bias effects are likely to be less sensitive to crossmodal SOA. Attentional cueing, however, facilitates perception when the cue precedes the target stimulus, but when the cue-target SOA increases above 300 ms the effect might reverse to an impediment, which is known as inhibition of return because it may reflect that the facilitating effect of attention has passed while attention is inhibited from returning to the same object. Reduction of temporal uncertainty works only when the cue is informative of the time of occurrence of the target and will not work when the cue-target SOA varies unpredictably. Therefore, we should be able to distinguish between perceptual and attentional effects by varying the crossmodal SOA.

2. Experiment 1

2.1. Methods

2.1.1. Participants

Twelve observers (seven females, mean age 26.0 years) participated in the experiments. All had normal hearing and normal or corrected to normal vision according to self report.

2.1.2. Stimuli and task – Experiment 1

The stimuli were presented in a 2IFC procedure. In one interval, the luminance either increased or decreased; in the other interval it remained constant. The order of the two intervals varied pseudorandomly and was counterbalanced across all trials for each observer. The sound was the same in both intervals. After the end of the second interval, the observer reported whether the luminance change had been in the first or the second interval. We call this the detection task. Then the observer was asked to report whether the

luminance change had been an increase or decrease. We call this the identification task.

The visual stimulus was a square of 1.0 degree visual angle. At the start of a trial, the square was filled with a mask of uniform random noise. After 500 ms the first interval started and the mask was replaced by a uniform square with a luminance that matched the mean luminance of the preceding mask. After 1000 ms the luminance level increased, decreased or remained constant. The square maintained this luminance level for an additional 1000 ms. This concluded the first interval. The mask was again shown for 500 ms and a second, similar interval followed. After the second interval, the mask reappeared.

The auditory stimulus consisted of samples from a uniform pseudorandom distribution. It started 500 ms after the interval begun – i.e. at the time that the mask was replaced by a uniform square. The sound lasted for 2000 ms – i.e. until the uniform square was replaced by a mask. A linear ramp of 10 ms was applied to the onset and offset of the sound to avoid very salient transients. The initial sound intensity of the sound was either 60 dB(A) or 75 dB(A). At 1000 ms after the sound commenced – i.e. at the time where the luminance might change, the sound intensity increased abruptly from 60 dB(A) to 75 dB(A), decreased abruptly from 75 dB(A) to 60 dB(A) or remained constant. There was a constant background noise at 38 dB(A) from computer ventilation. The timing of the audiovisual stimuli delivery occurred within 2 ms from the times reported here as measured using an oscilloscope connected to a photo diode and the sound card of the stimulus delivery computer.

2.2. Procedure

The experiment consisted of three parts: training, threshold estimation and threshold testing.

In the training part, the magnitude of the relative luminance change was 10%. The luminance either increased from 5.2 cd/m² to 5.7 cd/m² or decreased from 5.2 cd/m² to 4.6 cd/m². The sound intensity increased, decreased or remained constant. The type of luminance and sound intensity change varied pseudorandomly. The training session continued until the observer had responded correctly on nine out of ten consecutive trials in both the detection and the identification task. All observers completed the training part within a few minutes.

In the threshold estimation part, the magnitude of the relative luminance change varied according to two randomly intertwined adaptive staircases (accelerated stochastic estimation; Treutwein (1995)) set to adjust the proportion correct to 75%. This was done by setting the luminance change to the minimum change that the 8-bit graphics card could produce and then adjusting the baseline luminance in order to reach the threshold magnitude of the relative luminance change. One staircase adapted to the observer's threshold for detecting a luminance increase and the other for detecting a luminance decrease. The staircases continued until both had reached 20 reversals. The sound intensity was constant at 75 dB(A) during the threshold estimation trials.

In the threshold testing part, the magnitude of the relative luminance change was set to the mean of the detection thresholds for detecting a luminance increase and a luminance decrease. Two factors characterized a trial. The luminance change could either be an increase or a decrease and the sound intensity could increase, decrease or remain constant. This part thus contained $2 \times 3 = 6$ trial types which are depicted in Fig. 1. Each trial type was presented 28 times in pseudorandom order totaling $6 \times 28 = 168$ trials. The presentation was divided into eight blocks separated by a 10 s mandatory pause, which the observer was free to extend. The luminance change occurred in the first interval in half the trials and in the second interval in the other half of the trials. When the

sound intensity was constant, it was 65 dB(A) on half the trials and 70 dB(A) on the other half of the trials but observers' responses were pooled across these two stimulus types which were thus not distinguished in the following analysis.

The participants were instructed to report only what they saw and explicitly told not to base their response on the sound at all. They were also informed that the sound intensity change occurred at the same time as the luminance change and that the sound intensity was uninformative both of which interval contained the luminance change and whether the luminance had increased or decreased.

2.3. Experimental setup

The visual stimuli were presented on a Sony GDM F520 CRT monitor with a vertical refresh rate of 60 Hz and a resolution of 1024 by 768 pixels. The sound was presented through two balanced Bose Companion 2 desktop speakers placed to the sides of the monitor. Throughout the experiment the observer's head rested in a chin rest at 57 cm from the screen.

2.4. Results

Statistical tests were based on repeated measures ANOVA of the proportion of correct responses. Greenhouse-Geisser corrected p -values were also computed but as the correction did not change our conclusions uncorrected p -values and degrees of freedom are reported.

The mean proportion correct in the detection task is plotted in Fig. 3. The interaction between luminance change and sound intensity was not significant ($F(2,11) = 0.3$, $p > 0.7$) whereas the main effects of sound intensity ($F(2,11) = 4.7$, $p < 0.02$) and luminance ($F(1,11) = 5.3$, $p < 0.04$) were. The main effect of luminance is probably due the use of the average of the thresholds for luminance increase and decrease as the visual stimulus. More interestingly the lack of a significant interaction shows that a congruent and an incongruent sound intensity change had similar effects on visual detection sensitivity. To further test the effect of sound intensity, we calculated the difference in proportion correct between the conditions when the sound intensity changed and the condition when it remained constant. We found no significant interactions or main effects ($F(1,11) = 1.6$, $p > 0.2$) for all effects showing that the effect of a change in sound intensity did not depend on the direction of the luminance and sound intensity changes.

The mean proportion correct in the identification task is plotted in Fig. 3. The interaction between luminance change and sound

intensity was nearly significant ($F(2,11) = 3.4$, $p < 0.06$) whereas the main effects of luminance ($F(1,11) = 2.7$, $p > 0.1$) and sound intensity ($F(1,11) = 0.9$, $p > 0.4$) were not. Again we also analyzed the difference in proportion correct between the conditions when the sound intensity changed and the condition when it remained constant. The interaction between luminance change and sound intensity change was also nearly significant ($F(2,11) = 4.7$, $p < 0.06$) for the difference in proportion correct while the main effects of luminance ($F(1,11) = 1.2$, $p > 0.2$) and sound intensity ($F(1,11) = 0.2$, $p > 0.6$) were not.

To further investigate the different effects of sound intensity change in the detection and identification tasks, we conducted a between participants correlation analysis of the difference in proportion correct between the conditions when the sound intensity changed and the condition when it remained constant. In Fig. 4, the difference in proportion correct in the identification task is plotted against the difference in proportion correct in the detection task for each of the four conditions in which the sound intensity changed. We found a tendency for the increase in proportion correct due to sound intensity change to correlate between participants, when sound intensity and luminance increased. However, the effect was non-significant ($p < 0.03$) at a level of $0.05/4 = 0.013$ Bonferroni corrected for multiple comparisons. For the three other conditions, the correlations were non-significant ($p > 0.4$).

The identification task was performed in a yes/no-paradigm and thus prone to response bias. Therefore, we calculated the maximum proportion correct, which is the theoretical proportion correct that observers would obtain in the identification task if they used an unbiased criterion. It is thus a measure of the discriminability of the luminance increase and decrease. The mean \pm standard deviation maximum proportion correct for when the sound intensity increased, decreased or remained constant was 0.81 ± 0.12 , 0.78 ± 0.14 and 0.77 ± 0.10 respectively with no significant difference ($F(2,22) = 0.91$, $p > 0.4$) between the three conditions.

3. Experiment 2

3.1. Methods

Out of the 13 observers of Experiment 2 (eight females, mean age 26.2 years), 12 already participated in Experiment 1. All had normal hearing and normal or corrected to normal vision according to self report. The experimental setup and the stimuli were the same as in Experiment 1 with the exception that the sound inten-

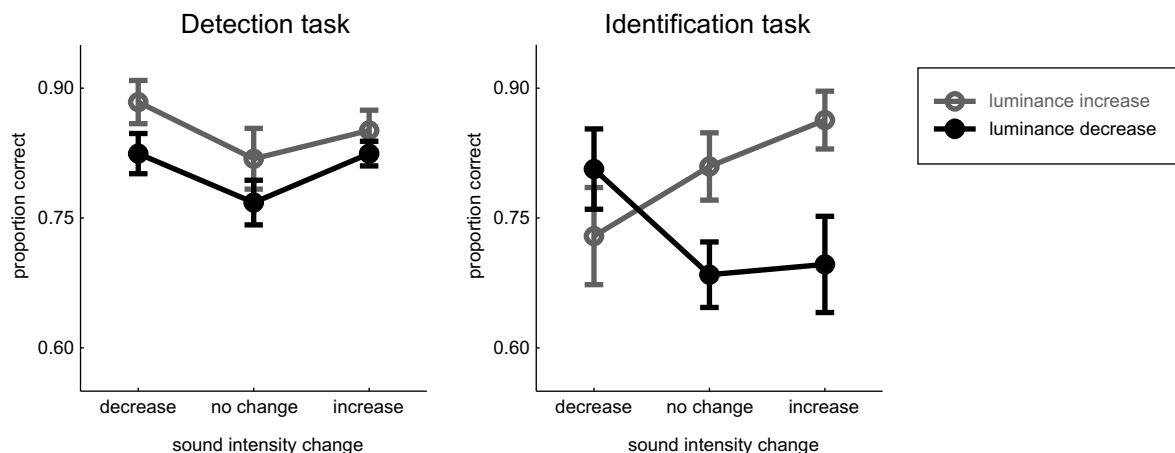


Fig. 3. Proportion correct in the detection and identification task of Experiment 1. Error bars represent the standard error of the mean.

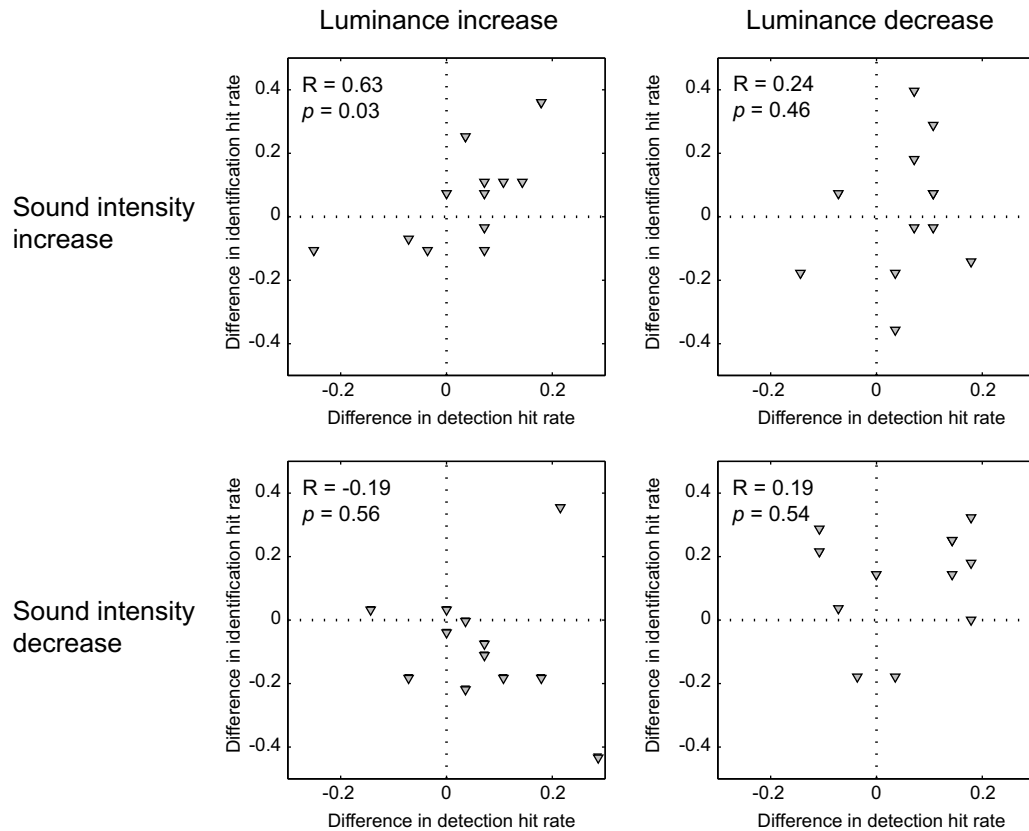


Fig. 4. Differences in proportion correct between the conditions when the sound intensity changed and the condition when it remained constant. The differences in the identification task are plotted against the differences in the detection task for all participants. Correlation coefficients, R , and p -values are given in the top left corner of each graph.

sity and luminance never decreased and that the sound intensity increase occurred at a latency of -400 , -150 , -75 , 0 , 75 , 150 or 400 ms with respect to the luminance increase. The observers only performed the detection task and not the identification task.

3.2. Results

The mean proportion correct as a function of audiovisual stimulus onset asynchrony (SOA) is plotted in Fig. 5. We found a signif-

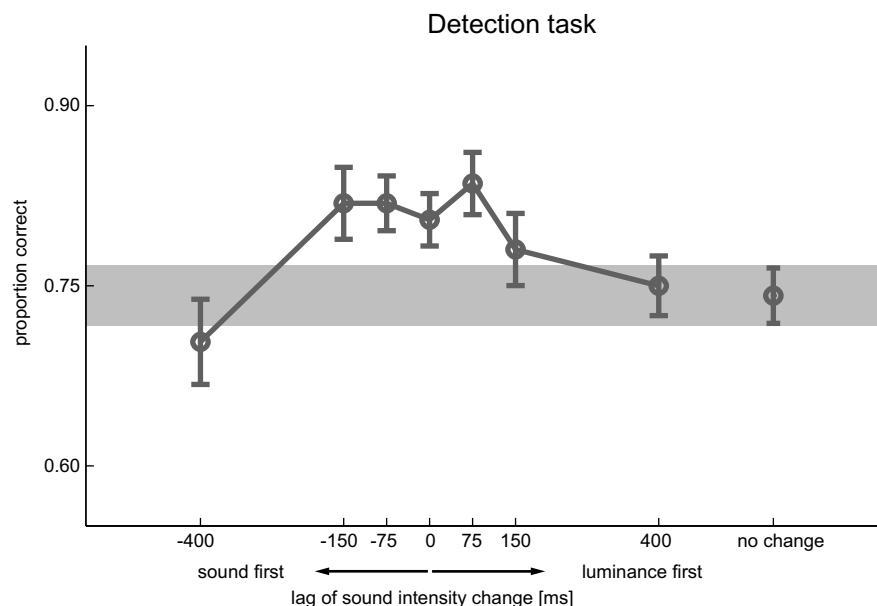


Fig. 5. Proportion correct in the detection task of Experiment 2 plotted against the lag between the sound intensity and luminance changes. Negative lags correspond to the sound intensity change preceding the luminance change. Error bars represent the standard error of the mean.

icant main effect of sound on proportion correct across the eight experimental conditions ($F(7,84) = 4.141, p < 0.001$). To investigate the effect further we compared each of the seven audiovisual SOAs with the condition where the sound intensity was constant. We found significant effects at latencies $-150, -75, 0$ and 75 ms ($F(1,12) > 4.9, p < 0.05$ for all conditions) but not at other latencies ($F(1,12) < 0.9, p > 0.3$ for all three conditions).

3.3. General discussion

In both experiments, we found that a sound intensity change facilitated detection of a luminance change. The effect occurred even though the change in sound intensity was irrelevant to the luminance change detection task and we therefore ascribe it to auxiliary integration.

The facilitation did not depend on whether sound intensity and luminance changed in the same direction in that there was no interaction between the effects of sound intensity and luminance change on observers' responses. Crossmodal interactions was thus not between signed internal representations of sustained loudness and brightness but between sound intensity and luminance change *per se*, i.e. between unsigned signal transients. This matches the predictions of the transient and attention hypotheses as depicted in Fig. 2A and D, respectively.

In the identification task, in which performance could only be based on a signed internal representation of luminance, we found a near significant interaction between change in luminance and sound intensity. In itself, this effect matches the predictions of the loudness/brightness and bias hypotheses as depicted in Fig. 2B and C, respectively. However, had it been a true perceptual effect as the loudness/brightness hypothesis propose, it would mean that a sound intensity change would decrease the magnitude of an incongruent luminance change and thus make it harder to detect but we did not find this in the detection task. We therefore ascribe the interaction effect to a response bias. Furthermore, we did not find any main effect of sound intensity showing that there was no effect of sound intensity on luminance sensitivity in the identification task even though this effect was strong in the detection task. Another way to look at this is to calculate the maximum proportion correct, which is an unbiased measure of observers ability to discriminate between a luminance increase and a decrease. We found that the maximum proportion correct was not affected by sound intensity. Remarkably, this means that even though a sound intensity change made observers better at seeing whether something happened it did not make them better at seeing what it was. This is in disagreement with the attention hypothesis, which predicts a general increase in visual sensitivity affecting performance in both the detection and identification task (cf. Fig. 2D). It is however, in good agreement with the transient hypothesis.

The dissociation of the effects of sound intensity change in the detection and identification tasks was also supported by the correlation analysis in which we found no significant correlation of the effects across participants. This lends further support to the notion that the effect of sound intensity in the detection task was due to perceptual integration while the effect of sound intensity in the identification task was due to a change in response bias. However, strictly speaking, we cannot exclude the possibility that the effect of sound in the identification task was, at least partially, perceptual. If performance in the detection task was based exclusively on unsigned stimulus transients while performance in the identification task was based exclusively on signed sustained intensity levels, perceptual integration could have occurred for both attributes independently. In other words, observers would have ignored their internal representation of sustained luminance levels when deciding in the detection task. Even though we find this explanation unlikely such an possibility has been put forward to

explain the dissociation between temporal order judgments and perceived simultaneity (Vatakis, Navarra, Soto-Faraco, & Spence, 2008).

In summary, from Experiment 1, we conclude that auxiliary audiovisual integration occurred between unsigned intensity transients but not between signed intensity levels, i.e. between loudness and brightness. Also, sound intensity change biased observers' responses in the identification task. This means that only a combination of the transient and bias hypotheses whose predictions are depicted in Fig. 2A and C, respectively, can explain our findings.

In Experiment 2, we found that sound induced facilitation of visual detection sensitivity persisted when the sound intensity change preceded the luminance change by as much as 150 ms and also when it followed the luminance change by as much as 75 ms. Two mechanisms are needed to explain these results. First, a temporal window for audiovisual integration of about 100 ms explains that the crossmodal facilitation tolerated audiovisual SOAs of ± 75 ms. Second, attentional pre-cueing enhances perception when the cue precedes the target with as much as 300 ms. This can explain why the sound intensity change facilitated visual detection sensitivity when the sound intensity change preceded the luminance change by 150 ms; in this case it might not have been a matter of multisensory integration but of the sound intensity change acting as a cue to attention that the luminance change was about to happen. This also explains the asymmetrical lack of effect when the sound intensity change *followed* the luminance change by 150 ms; at this audiovisual SOA, neither multisensory integration nor attentional cueing would be at work. Also, the non-significant tendency for sensitivity to decrease when the cue preceded the target with 400 ms might be ascribed to inhibition of return. We emphasize that there is no discrepancy between adopting an attentional explanation when the sound intensity change precedes the luminance change in Experiment 2 while ruling out attentional effects when the sound intensity and luminance change were simultaneous in Experiment 1 because attentional effects depend crucially on cue-target latency.

In Experiment 2, the audiovisual SOA varied pseudorandomly and the sound intensity change was therefore uninformative of the onset of the luminance change. Thus, the sound intensity change did not reduce the temporal uncertainty of the onset of the luminance change. Yet, we found that the sound intensity change still caused an increase in visual detection sensitivity. This confirms that enhanced sensitivity is not due to reduction of temporal uncertainty and thus further support that the observed effects reflect perceptual integration of audiovisual stimulus transients.

Audiovisual signal detection has also been studied extensively physiologically (Stein & Meredith, 1993). In a series of studies, Stein and coworkers described how some neurons in the Superior Colliculus have a multisensory receptive field in that they respond weakly to auditory or visual stimulation alone but strongly to audiovisual stimuli. However, as Schnupp et al. pointed out, these neural effects could reflect both combinatory and auxiliary integration. Still, a study by Frassinetti, Bolognini, Bottari, Bonora, & Ladavas (2005) supports the idea that the Superior Colliculus underlies auxiliary audiovisual integration in signal detection. They showed that an irrelevant sound can also improve luminance detection sensitivity in the blind hemifield of hemianopes. They deduced that since the cause of the hemianope's blindness is cortical damage this finding points to a subcortical system such as the Superior Colliculus mediating audiovisual integration. They proposed to test this finding by using visual stimuli with a wavelength only visible to S-cones because the Superior Colliculus does not receive input from S-cones. Consequently, in a preliminary report, Maravita, Savazzi, Bricolo, Penati, & Marzi (2005) showed that sound does

not shorten reaction times to S-cone stimuli more than could be expected if audition and vision were independent processes supporting the notion that auxiliary audiovisual integration occurs in a part of the visual system, such as the Superior Colliculus or the magnocellular pathway, that does not receive information from S-cones (Sumner, Adamjee, & Mollon, 2002).

In addition, Superior Colliculus neurons' responses saturate at low contrast, which indicates that it would be involved in signal detection. Also, besides being insensitive to S-cone output, the Superior Colliculus responds mostly to signal transients and hardly to sustained stimuli at all (Schneider & Kastner, 2005). This makes the Superior Colliculus suitable for change detection and accordingly multisensory integration in the Superior Colliculus is likely to be important for orientation responses to weak stimuli (Jiang, Jiang, & Stein, 2002). Our finding of audiovisual integration of intensity transients but not of sustained intensity levels are thus in excellent agreement with the integration occurring in the Superior Colliculus. Furthermore, the Superior Colliculus responds strongly to audiovisual stimuli only if the auditory and visual stimulus coincide within a temporal window of 100 ms. This temporal window also applies to the sound induced perceptual enhancement of visual transients that we report here. We therefore suggest that our results reveal that the perceptual consequence of the well-established multisensory integration in the Superior Colliculus is an integration of intensity transients.

Acknowledgments

This study was supported by a chaire d'excellence from the French Ministry of Research and a grant from the Danish Council for Strategic Research.

References

- Andersen, T. S., Tiippana, K., & Sams, M. (2004). Factors influencing audiovisual fission and fusion illusions. *Brain Research Cognitive Brain Research*, 21(3), 301–308.
- Frassinetti, F., Bolognini, N., Bottari, D., Bonora, A., & Ladavas, E. (2005). Audiovisual integration in patients with visual deficit. *Journal of Cognitive Neuroscience*, 17(9), 1442–1452.
- Frassinetti, F., Bolognini, N., & Ladavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Experimental Brain Research*, 147(3), 332–343.
- Green, D. M., & Swets, J. A. (1966). *Signal Detection and Psychophysics*. New York: John Wiley & Sons.
- Jiang, W., Jiang, H., & Stein, B. E. (2002). Two corticotectal areas facilitate multisensory orientation behavior. *Journal of Cognitive Neuroscience*, 14(8), 1240–1255.
- Macaluso, E., Frith, C. D., & Driver, J. (2000). Modulation of human visual cortex by crossmodal spatial attention. *Science*, 289(5482), 1206–1208.
- Maravita, A., Savazzi, S., Bricolo, E., Penati, V., & Marzi, C. A. (2005). *Role of Superior Colliculus in Audio-Visual Redundancy Gain*. *International Multisensory Research Forum*. Italy: Rovereto.
- McDonald, J. J., Teder-Salejari, W. A., & Hillyard, S. A. (2000). Involuntary orienting to sound improves visual perception. *Nature*, 407(6806), 906–908.
- McDonald, J. J., Teder-Salejari, W. A., & Ward, L. M. (2001). Multisensory integration and crossmodal attention effects in the human brain. *Science*, 292(5523), 1791.
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746–748.
- Meredith, M. A., Nemitz, J. W., & Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *Journal of Neuroscience*, 7(10), 3215–3229.
- Nobre, A. C., & O'Reilly, J. (2004). Time is of the essence. *Trends in Cognitive Sciences*, 8(9), 387–389.
- Odgaard, E. C., Arie, Y., & Marks, L. E. (2003). Cross-modal enhancement of perceived brightness: Sensory interaction versus response bias. *Perception & Psychophysics*, 65(1), 123–132.
- Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America, A*, 2(9), 1508–1532.
- Posner, M. I. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology*, 32(1), 3–25.
- Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology*, 109(2), 160–174.
- Recanzone, G. H. (2003). Auditory influences on visual temporal rate perception. *Journal of Neurophysiology*, 89(2), 1078–1093.
- Schneider, K. A., & Kastner, S. (2005). Visual responses of the human superior colliculus: A high-resolution functional magnetic resonance imaging study. *Journal of Neurophysiology*, 94(4), 2491–2503.
- Schnupp, J. W., Dawe, K. L., & Pollack, G. L. (2005). The detection of multisensory stimuli in an orthogonal sensory space. *Experimental Brain Research*, 162(2), 181–190.
- Sekuler, R., Sekuler, A. B., & Lau, R. (1997). Sound alters visual motion perception. *Nature*, 385(6614), 308.
- Shams, L., Kamitani, Y., & Shimojo, S. (2000). Illusions. What you see is what you hear. *Nature*, 408(6814), 788.
- Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Brain Research. Cognitive Brain Research*, 14(1), 147–152.
- Spence, C., & Driver, J. (1997). Audiovisual links in exogenous covert spatial orienting. *Perception & Psychophysics*, 59(1), 1–22.
- Stein, B. E., London, N., Wilkinson, L. K., & Price, D. D. (1996). Enhancement of Perceived Visual Intensity by Auditory Stimuli: A Psychophysical Analysis. *Journal of Cognitive Neuroscience*, 8, 497–506.
- Stein, B. E., & Meredith, M. A. (1993). *The Merging of the Senses*. MIT Press.
- Sumner, P., Adamjee, T., & Mollon, J. D. (2002). Signals invisible to the collicular and magnocellular pathways can capture visual attention. *Current Biology*, 12(15), 1312–1316.
- Treutwein, B. (1995). Adaptive Psychophysical Procedures. *Vision Research*, 35(17), 2503–2522.
- Vatakis, A., Navarra, J., Soto-Faraco, S., & Spence, C. (2008). Audiovisual temporal adaptation of speech: Temporal order versus simultaneity judgments. *Experimental Brain Research*, 185(3), 521–529.
- Warren, D. H. (1979). Spatial localization under conflict conditions: Is there a single explanation? *Perception*, 8(3), 323–337.